**Recomendaciones SaludAlpes**

**Link del repositorio:**

**La presentación se encuentra en los entregables como Presentación SaludAlpes.pptx**

**El dashboard se encuentra en el repositorio.**

1. Tratamiento de los datos:

Para comenzar con la construcción del modelo decidimos eliminar las siguientes columnas porque consideramos que no eran relevantes para el diagnóstico de la diabetes en los pacientes:

* Color de cabello.
* Ciudad.

Posteriormente, procedimos a eliminar las filas con valores nulos e inconsistentes en las columnas: “Insulin”, “Skin thickness”, “Glucose”, “BMI” y edad. La anterior decisión se tomó debido a que resultó ser de suma importancia alimentar al modelo con datos consistentes ya que este ayudará al diagnóstico de los pacientes.

1. Construcción de los modelos:

**KNN** (Nicolás Segura): Para la construcción de este modelo se realizó la previa normalización de los datos con el fin de que las distancias fueran equiparables. Además, se utilizaron los siguientes hiperparámetros validados con el algoritmo k-fold-cross validation: **Minkowski**, **p**=2, **vecinos** =5.

**Árbol de decisión** (Nicolás Segura): Debido a los cambios realizados con anterioridad al conjunto de datos, no hubo necesidad de hacer uno adicional. Los hiperparámetros utilizados para la construcción del modelo también se validaron a través de k-fold-cross validation y fueron: **Gini**, **mm**=5, **profundidad**=4.

1. Comparación y selección de los modelos:

Una vez obtenidos los resultados de la precisión y el puntaje f1 de cada uno de los modelos decidimos seleccionar el modelo de **Árbol de decisión**, ya que a pesar de que contaba con un 2% menos de precisión que el de **Regresión lineal**, este contaba con un mejor puntaje f1, lo que en este caso resulta de mucha más importancia ya que dicho indicador permite saber qué tan bueno es el modelo para evitar los falsos negativos; los cuales resultan ser de vital importancia para SaludAlpes para evitar dar diagnósticos que puedan perjudicar la salud del paciente por la omisión de un tratamiento temprano.